One of the first modules in this course presented some useful ways to think about causes. And one of the concepts that was stressed is that multiple factors often work together to cause a particular outcome. This is certainly true of most unwanted health outcomes.

The last module on confounding discussed the confusion that other causes introduce when we are trying to quantify the effect of just one component of cause. For example, we all know that there are multiple risk factors that can contribute to the occurrence of coronary heart disease. While acknowledging this, we might want to conduct research to get a precise measure of the impact of physical activity in preventing heart disease. And if this is our goal, all of the other component causes are just noise that make it difficult to tease out the specific role of physical activity.

We learned in the last module that stratified analysis is a method for obtaining a measure of association that is adjusted for one or more confounding factors. However, stratification is only practical for adjusting for one or two factors at a time, even though we know that diseases like coronary heart disease are multi-factorial and have many contributing component causes.

This brings me to two key points. The first is that to truly understand the potential impact of exercise in preventing heart disease, I would ideally measure its impact while controlling simultaneously for as many other causes as possible.

The second point is that to achieve real insight about the causes of heart disease, it would be very desirable to have an unconfounded model that enabled us to see the relative contributions of each of the many component causes that contribute to heart disease. In fact, it is possible to work towards this degree of understanding using multiple variable regression analysis, a form of analysis that enables us to construct mathematic models that describe the association between a disease outcome and its many component causes while adjusting for confounding.

Our starting point will be simple linear regression, which we discussed in an earlier module. A simple regression model is a way of describing mathematically the relationship between a single independent or predictor variable and a dependent or outcome variable. We will then build on this to give you a tool for dealing with multiple confounding factors simultaneously using R.